DOCUMENT RESUME

ED 395 938 TM 025 005

AUTHOR Kirsch, Irwin S.; Mosenthal, Peter B.
TITLE Understanding Document Literacy: Variables Underlying the Performance of Young Adults.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-88-62
PUB DATE Nov 88
NOTE 72p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Blacks; *Cognitive Processes; *Educational Attainment; Ethnic Groups; Hispanic Americans; *Literacy; Models; National Surveys; *Performance Factors; *Reading Achievement; Scoring; Whites; *Young Adults
IDENTIFIERS *Document Literacy Skills; National Assessment of Educational Progress

ABSTRACT

Critical variables that underlie the performance of a national sample of young adults on a diverse set of document literacy tasks were identified. The final sample was 3,618 adults. The identification of these variables provides an important first step toward building a theoretical model that would systematically account for the constructs of document processing. The 61 tasks and their associated documents that make up the document scale of the National Assessment of Educational Progress (NAEP) Young Adult Literacy assessment were parsed using a specially devised grammar. Based on the parsings, variables were identified to account for the probability of success for the total population and for major subgroups of interest. The identified variables accounted for 89% of the variance for the total population. Among racial and ethnic groups, these variables accounted for 89% of the variance for White, 81% for Black, and 87% for Hispanic young adults. Among levels of education, these variables accounted for 56% of the variance for young adults with 0 to 8 years of schooling, 81% for young adults with 9 to 12 years of schooling, 88% for young adults with high-school degrees, and 84% for young adults with post-high-school degrees. Findings are discussed in terms of the need to provide a more general framework for describing, comparing, and researching documents. Five appendixes discuss parsing and scoring. (Contains 2 figures, 3 tables, 1 appendix table, and 70 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 395 938

RR-88-62

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

UNDERSTANDING DOCUMENT LITERACY: VARIABLES UNDERLYING THE PERFORMANCE OF YOUNG ADULTS

Irwin S. Kirsch
Peter B. Mosenthal

BEST COPY AVAILABLE

Educational Testing Service
Princeton, New Jersey
November 1988

TM025005

Understanding Document Literacy:

Variables Underlying the Performance of Young Adults

Irwin S. Kirsch

Educational Testing Service

Peter B. Mosenthal

Syracuse University

Running Head: Understanding Document Literacy

The data analyzed in this study were collected through the National Assessment of Educational Progress (NAEP) study of Young Adult Literacy. The work reported in this paper was supported by Educational Testing Service (ETS) as part of a larger effort to develop a learning and assessment system for adults in the area of literacy. Grateful acknowledgement is due to Claire Putala and Claudia Gentile for their assistance in conducting reliability studies of the grammar. We also thank Ann Jungeblut and Philip Oltman for their thoughtful comments on the manuscript.

Copyright © 1988. Educational Testing Service. All rights reserved.

Abstract

The purpose of this study was to identify a set of critical variables that underlie the performance of a national sample of young adults on a diverse set of document literacy tasks. The identification of these variables provides an important first step toward building a theoretical model that would systematically account for the constructs of document processing. With such a theoretical model, document designers and instructional-program developers could use their understanding of document-processing constructs in ways that could strategically address the production and processing of documents.

The sixty-one tasks (and their associated documents) that make up the document scale of the NAEP Young Adult Literacy assessment were parsed using a specially devised grammar. Based upon the parsings, variables were identified to account for the probability of success for the total population and for major subgroups of interest. The variables identified accounted for 89 percent of the variance for the total population of young adults. Among racial/ethnic groups, these variables accounted for 89 percent of the variance for White, 81 percent for Black, and 87 percent for Hispanic young adults. Among levels of education, these variables accounted for 56 percent of the variance for young adults with 0-8 years of schooling, 81 percent for young adults with 9 to 12 years of schooling, 88 percent for young adults with high-school degrees, and 84 percent for young adults with post-high-school degrees. The findings of this study are discussed in terms of the need to provide a more general framework for describing, comparing, and researching documents than has been the case in previous document studies.

Understanding Document Literacy:

Variables Underlying the Performance of Young Adults

Introduction

Using the database provided by the NAEP literacy assessment of young adults (Kirsch & Jungeblut, 1986), the current study sought to identify a set of critical variables that underlie young adults' ability to perform document literacy tasks. By identifying these variables for a wide range of documents and relating them to the performance of a nationally representative sample of 21- to 25-year olds, this study extends our understanding of what makes documents "simple" or "usable." Without a firmer theoretical framework than currently exists, there is no starting point for identifying the constructs underlying ability to use documents effectively (Kirsch & Guthrie, 1980). Knowledge of such constructs provides the theoretical basis for designing effective instructional programs (Messick, in press). Moreover, in the absence of construct specification, there can be no set of generalizable and empirically determined principles for designing simple and usable documents (Wright, 1978, 1988).

The introduction considers the following. First, to provide an overview on this topic, the importance and pervasiveness of documents in today's society are discussed. Next, some of the problems associated with producing and processing documents are considered. How document researchers have attempted to deal with these problems using the "redesign-test-redesign-and-test" approach is noted and the limitations of this paradigm are described.

In this paper we have adopted the definition of document literacy put forward by NAEP (Kirsch & Jungeblut, 1986). Document literacy involves the knowledge and skills needed to understand and use printed information occurring in a variety of non-prose formats. Non-prose formats include linguistic structures that are not organized in paragraph form. As such, non-prose formats consist of the following: forms, tables, charts, graphs, signs/labels, indexes, lists, schematics, and catalogues.

The Importance of Documents

As Burch and Grudnitski (1986) have noted, there are few actions that can take place in today's society that do not require the use of documents. For example, people apply for jobs by filling out a job application. Patients are admitted to hospitals by completing admissions forms. Students enroll in their courses by filling out a specific registration form. Building construction can not proceed without completing forms for a building permit. Indeed, couples who wish to legalize their coexistence complete the necessary forms to obtain a marriage license.

In addition, documents can be used to inform people's actions and decisions. For example, to refine our diet, we may turn to an almanac table that lists the various food sources of Vitamin E. We may consult our benefits table to see if our employer covers a certain type of disability. As business forecasters, we may refer to a bar graph that summarizes the sales trend of a company by season over a three year period. We may also refer to indexes to assist us in locating specific information to answer a question. To get the rust removed from our car, we may consult the Yellow Pages listings for the

names and numbers of the different autobody repair shops. Finally, to complete our income tax, we may consult our Federal Income Tax table.

Documents also provide records of actions taken, decisions made, and agreements reached. For example, as a dispatcher, we may detail the number and times different limousines are sent to the airport. As investors, we fill out deposit slips to place money into our mutual fund accounts. As sales clerks, we enter buyers' credit card numbers on a bill of purchase. As beginning professors, we may sign a job contract that represents a mutually agreed upon set of conditions between our university employer and ourselves. And, we may draw up a will specifying the desired disposition of our assets at death.

All these examples illustrate the importance of documents in our daily lives. In addition to these uses, numerous other functions of documents have been identified (Bassett, Goodman, & Fosegan, 1981; Wright, 1988). For example, Bassett, Goodman, and Fosegan (1981) have identified over twenty basic document functions: among these, documents grant authority, help ensure control, provide audit trails, and notify people of changes in status.

The Pervasiveness of Documents

In addition to serving important functions, the number of paper documents tends to increase significantly each year. Even today in organizations relying heavily on advanced information technology, paper documents are still the primary means used to get data into the system (Burch & Grudnitski, 1986). Although the actual number of documents produced each year is not known, several studies (e.g., Rayner, 1982; United Kingdom, 1982;

Waller, 1984) have illustrated the pervasiveness of one type of document, namely forms, within government.

In a survey of British government forms, Rayner (1982) estimated that the number of different external forms--those issued to the public or to business organizations--was about 38,000 and that the number of internal administrative forms was about double this assessment. He further estimated the total number of **different** government forms to be well over 100,000. In another study, the Associated Press (Miller, 1984) has estimated that in the mid 1970s, the United States government issued some 98,000 different kinds of forms per year, and received over 50 million responses. In the same time period, the US Internal Revenue Service alone sent out over 3,500 different forms per year.

Given the increase in information necessary to maintain government, military, industry, and other institutional establishments, we can be assured that the number of documents issued will continue to increase significantly, if not prohibitively, in the years to come (Waller, 1984).

The pervasiveness of documents within our society is reflected not only by their actual numbers, but also by the amount of time spent reading or using these various materials. One study (Guthrie, Seifert, & Kirsch, 1986) randomly sampled wage earners from over a 100 households within a community of 6,000 in an effort to measure adult readership. Volume, as measured by minutes of reading per day, was estimated across contents that included: news/business, society/science, recreation/sports, fiction/viewpoint, reference, and brief documents. The amount of time spent reading these

contents was estimated for occupational groups, level of education, and gender. For example, males reported reading various contents for a total of 161 minutes each day compared with 221 minutes reported by females.

Among occupational groups, managers/professionals reported the highest volume of reading -- an average of 269 minutes each day. They were followed by clerical workers who reported reading for an average of 228 minutes per day. Skilled workers reported spending 160 minutes each day reading various contents and unskilled workers an average of 98 minutes each day.

Within each of the groups, the largest percentage of time spent reading each day was with the content of brief documents. For example, among females, 97 of the 221 minutes, or 44 percent of the reported daily reading involved brief documents. In contrast, the next highest content reported among females was society/science -- an average of 36 minutes which accounted for only 16 percent of daily reading volume. Similarly, among males, 68 minutes or 42 percent of daily reading volume was associated with the content of brief documents. The next highest volume involved the content of news/business which was read for an average of 30 minutes or 19 percent of daily volume. Similar patterns of results indicating high percentage volume for brief documents and much lower for other contents were observed for each occupational group and at all levels of education.

In a separate study, Kirsch & Guthrie (1984a) examined the reading practices of adults in one high-technology corporation. One aspect of that study involved estimating reading volume for work and leisure across the same six content areas noted above. In addition to estimating reading associated

with work and leisure settings, results were reported by occupation and education as well. Again, regardless of level of education or type of job held in the corporation, the largest percentage of time spent reading each day involved brief documents.

This study also yielded evidence that reading brief documents occurs primarily at work. In contrast, the other content areas are associated with leisure reading activities. Whereas these two studies suggest distinctions between types of reading engaged in for work and leisure, other research has pointed to distinctions between reading done for school and reading associated with other adult contexts (Mikulecky, 1982; Sticht, 1977; Venezky, 1982).

Problems Associated with Producing and Processing Documents

Effective document performance not only has consequences for individuals; it has significant consequences for organizations as well. In short, for organizations to function effectively, they must maintain accurate, timely, and relevant information (Burch & Grudnitski, 1986). The vehicle for securing such information is documents.

Different groups of document users make up an interactive chain that define the effectiveness of document performance within an organization. These groups include: (1) interpreters of information, such as managers, who use document-generated information for controlling, planning, and decision making; (2) providers of information, such as accountants, tax payers, and bill recipients; (3) information-support personnel, such as secretaries, programmers, administrators, and systems analysts; and (4) designers and producers of documents.

In an attempt to define document effectiveness, researchers (Athey & Zmud, 1986; Duffy, 1981) typically distinguish between the costs of processing documents (e.g., a form) and the costs of producing a document. Most investigators (e.g., Waller, 1984) have found that the principal costs lie in the processing of forms, with the costs of processing forms typically exceeding the costs of producing forms by a factor of 2 to 3 (Rayner, 1982).

A major factor contributing to the processing costs of documents is the number of mistakes that providers of information make in completing a document. Errors made on forms lead to greater processing time; this time is often compounded by the problem of having to return documents for correction. Although few studies have explicitly looked at the impact of document error frequency on company costs, studies have identified the large occurrence of errors in many widely used documents. For example, Waller (1984) reported that the old British P1 income tax form had an error rate of 84 percent. The United States 1988 income tax form is expected to have an error rate of 57 percent (CNN News, 1988).

Additional factors contributing to an organization's information processing costs include: (1) decision makers' misinterpretation of document information (Odell & Goswami, 1981) and (2) the tendency of information-support personnel to incorrectly re-enter information from a paper document to a computer database (Athey & Zmud, 1986).

In sum, the point is that for documents to be effective for an organization, they must be effective for each group in the chain of document producers and users.

The "Redesign-Test-Redesign-and-Test" Approach

The growing importance that documents are expected to play in our daily lives coupled with the problems identified in processing these materials has led government, military, and industry to consider several options. These include: (1) reducing the number of documents; (2) raising the level of literacy skills among individuals; and, (3) making documents more readable and usable (Department of Health and Social Security, 1983; Duffy, 1985; Felker & Rose, 1981; Hartley, 1985; Sticht, 1975). Of these three options, the one that has received increasing attention in the last ten years involves research aimed at designing more effective documents. This effort has been encouraged by the Plain English campaigns in the United States and England to simplify the language of documents and to make them more interpretable to users (Chapanis, 1965; Jereb, 1986; Shilling, 1981).

Attempts to simplify documents (e.g., Atwood, Baker, & Duffy, 1985; Charney, 1986; Firth, 1981; Janik, Hannah, Waney, Bond, & Hayes, 1981; Keller-Cohen, 1987; Wright, 1980a; Waller, 1984) have built upon Wright's (1979) suggested "redesign-test-redesign-and-test" approach. In this approach, a commonly used document is first identified. Next, using a set of criteria, the document is "simplified" or rendered more "usable." The criteria for rewriting documents are varied. They may be based on readability formula criteria (e.g., Atwood et al., 1985), user or designer intuition (e.g., Firth, 1981; Janik et al., 1981; Waller, 1984), general document design principles (Hartley, 1985; Landesman, 1981; Raines, 1980; Wright, 1981, 1984), or some theory or model of reading (e.g., Atwood et al., 1981; Wright, 1984).

Both versions of the document are then subjected to some processing measure, such as the speed and accuracy by which two groups of users complete the two versions respectively. A common evaluation measure is having users report aloud their processes as they attempt to complete each of the two versions (e.g., Charney, 1986; Janik et al., 1981; Schumacher & Waller, 1985). Document revision and testing continue until the modified document is shown to be more effective than the original, relative to a selected criterion. Researchers then conclude that the factors which significantly influence a document's readability or usability have been identified.

Limitations in Implementation of the "Redesign-Test-Redesign-and-Test" Approach

While the above paradigm has been useful in terms of improving individual documents, it lacks generalizability for several reasons. First, the subjects in document redesign-test studies often are not representative of the users for whom the document was originally designed or intended. As Firth (1981) has shown, people who actually use the documents may perform in ways that differ significantly from subjects in a study who would not normally use the document. Some studies have drawn their samples from actual users of the document. However, the number of subjects in these studies who are also actual users of the document is so limited as to raise questions concerning the generalizability of the findings (Wright, 1988).

Second, generalizability is constrained not only by the small number of subjects, but by the limited number of documents studied as well. Thus, because "redesign-test" studies typically use only one document, the findings

that hold from one study can not necessarily be generalized to studies that use different documents (Hartley & Trueman, 1985). As a consequence, the ways in which this approach has been implemented are subject to Clark's (1972) criticism that, because the document itself is a "fixed effect," the findings may be replicable for the particular document studied, but the results provide little or no information that can be generalized.

A third factor contributing to the lack of generalizability is the fact that no descriptive or explanatory grammar has been developed for use with documents, as opposed to expository prose (Fredericksen, 1975; Kintsch, 1977; Meyer, 1975). Such grammars serve as important heuristics for comparing the structures and content of various prose passages (de Beaugrande, 1981). Because no such heuristic exists for analyzing documents, it is not possible to compare and contrast systematically the structures and content of different documents. (Atwood, Baker, & Duffy, 1985; Bovair & Kieras, 1981; de Beaugrande, 1980; Kieras & Dechert, 1985).

Finally, researchers employing this "redesign-test" approach often use different administration and scoring procedures in testing a document's effectiveness from those found in the actual use of this document (Keller-Cohen, 1987; Wright, 1980b). This is problematic because different types of questions, directives, and scoring procedures produce different document user outcomes (Atwood, et al., 1985; Barnard, Wright, & Wilcox, 1979; Wright, 1988). Thus, if the conditions under which the subjects completed the document task are different from those under which real users normally

complete the document task, then the performance of the subjects is not a good predictor of the performance of the real users.

The design of the NAEP assessment addressed these limitations in the following manner: it provides a database for studying how a broad range of document configurations and different simulation tasks interact to influence document performance. Rather than using multiple choice exercises, the majority of document literacy tasks required the respondents to engage in those procedures actually associated with the materials. For example, respondents were directed to: follow a set of directions to travel from one location to another using a map; locate information using an index from a reference book; fill in a deposit slip; complete a check; determine eligibility from a table of employee benefits; and, fill out an order form taken from a catalogue. Moreover, the data base provides an opportunity to compare and contrast the structures and contents of an array of document types that could be used to establish a set of generalized principles for future examination.

Methods Underlying the NAEP Assessment

Subjects

The target population for the Young Adult Literacy Assessment consisted of young adults in the continental United States who, at the time of the assessment (April through September, 1985), resided in private households and who were between the ages of 21 and 25. The goal of the sample design was to achieve a projectable sample of this target population and to oversample Black

and Hispanic young adults at approximately double their normal rate so that reliable estimates of proficiency could be obtained. A total of 38,400 housing units in 800 locations were screened for eligible respondents. Of the 4,494 young adults who were selected for the assessment, interviews were completed with 3,618. This represents an assessment completion rate of 80.5 percent. An incentive of $15 was offered to each respondent for participating and completing the assessment. Additional details and considerations involving the sampling, weighting, and data collection activities can be found in Kirsch & Jungeblut (1986).

Materials

NAEP's Development of Simulation Tasks

In selecting materials and developing tasks for inclusion in the assessment, primary emphasis was placed on representing the broad range of literacy behaviors that people frequently encounter in occupational, social, and educational settings. To assist in determining the nature of such materials and tasks, lists of current objectives in competency-based adult programs, existing literacy measures, and studies of literacy in various contexts were reviewed (Kirsch & Jungeblut, 1986).

Based on this review, 12 categories of materials were identified. These included: sign/label, directions, memo/letter, form, table, graph, prose, index/reference, notice, schematic or diagram, advertisements, and bill/invoice. These categories were then crossed with five categories of use: knowledge, evaluation, specific information, social interaction, and

application. Use or purpose refers to why individuals might engage in a task, i.e., the type of information they need or are seeking. This is believed to influence both a person's strategy and cognitive operations in completing the task (Crandall, 1981; Kirsch & Guthrie, 1984b; Scribner & Cole, 1981; Sticht, 1978, 1982). The matrix that resulted from crossing materials with purpose provided the framework that was used to develop tasks. It should be noted that tasks were not developed to fill all cells of this matrix. To do so would have meant that many of the tasks would not have been representative of the kinds of tasks that adults normally encounter. In total, some 105 scorable tasks were developed and selected for inclusion in the assessment.

Organizing Tasks into Blocks

Because only 60 minutes of response time was allocated to the measurement of literacy skills, it was necessary to employ some form of item sampling procedure to ensure broad and representative coverage of content. A powerful variant of standard matrix sampling, called Balanced Incomplete Block (BIB) spiraling, was used. In BIB spiraling, as in standard matrix sampling, no respondent is administered all of the tasks in the assessment pool. However, unlike standard matrix sampling in which items are assembled into discrete booklets, BIB spiraling allowed for the estimation of relationships among all tasks in the pool through the unique linking of blocks. In the NAEP assessment, the total item pool was divided into seven blocks of tasks, with each block requiring approximately 17 minutes of administration time based on field trials. Each respondent completed one of seven assessment booklets,

BEST COPY AVAILABLE

each of which included three blocks of tasks, a set of seven core tasks, and a 30-minute background questionnaire.

One important outcome of this BIB spiraling design was that every task was taken by a randomly equivalent subsample of the total sample of respondents. In this assessment, approximately 1,500 individuals responded to each task, except for the core which was attempted by all respondents. This type of design ensured that reliable estimates of performance of the population as a whole, as well as of major subgroups of interest, could be derived for each task. That is to say, reliable estimates of the percentage of each subgroup of interest who responded correctly to each task were obtained (Kirsch & Jungeblut, 1986).

Procedures

Data Collection and Quality Control

Data collection activities were performed by Response Analysis Corporation (RAC) field staff over a six month period. Approximately 500 trained interviewers conducted individual assessments in the respondents' homes. Each interviewer received and studied training materials and conducted a practice interview. These materials were reviewed by RAC staff and consultations were provided to clarify any problems noted. Interviewers who were not highly experienced in interview procedures received additional training in sampling procedures, general interviewing techniques, and in administration of the assessment instruments. This training was conducted in person with area supervisors in 12 regional training sessions.

Each interviewer had responsibility for screening a selected set of households, for following explicit instructions, for selecting an eligible respondent within a household, and for conducting the assessment. In general, the interviewer acted as a neutral proctor throughout every individually-administered assessment. The interviewer guided the respondent through the assessment using standardized instructions contained in interview guides constructed for each of the seven assessment booklets.

To help assure that correct procedures were being followed, RAC's office staff reviewed all completed assessments. This involved: reviewing each of the listings made; coding each interview with respect to the booklet used and respondent characteristics; and reviewing each assessment to insure that a set of key questions from each phase of the assessment had been completed. In addition, 25 percent of each interviewer's completed assessments were subjected to a quality control check. This check involved contacting the selected respondents and verifying with them a selected set of questions.

Scoring

Most of the literacy tasks contained in this assessment involved open-ended responses. A scoring guide was developed for each. During a one-week period, eight individuals were trained by a supervisor to read and score the responses. The scoring guides were discussed during the training period and each trainee practiced on a selected set of actual responses. The scores assigned were discussed by the group, resulting in some clarification to the rules for assigning scores.

All of the open-ended tasks in the assessment were subject to a 20 percent reliability check involving a second reading by a second scorer. Scorer reliability was estimated on a weekly basis. Overall, the average percent agreement among the eight scorers for all open-ended tasks was 96 percent. However, reliabilities among scorers on any given task, ranged from 86 to 100 percent.

Methods Underlying the Current Study

Materials

The Document Tasks

As noted earlier, the NAEP assessment contained some 100 tasks that simulated the levels and types of processing associated with using printed materials in our society. Based on theoretical considerations, NAEP chose to represent these tasks in terms of three categories or families of tasks: prose literacy, quantitative literacy, and document literacy. Prose literacy tasks required the reader to demonstrate knowledge and skills associated with understanding and using information from texts that included editorials, news stories, poems, and the like. Quantitative tasks required respondents to perform different arithmetical operations, either alone or sequentially, with information that was embedded in either prose or document formats. Document tasks involved the knowledge and skills required to locate and use printed information that was not in prose format, such as tables, charts, indexes, forms, maps, schedules, and bills. In total, the 61 document tasks (involving

37 different materials) that comprised this scale provided the data for the current study.

These 37 materials, along with the corresponding questions and directives, were analyzed relative to their structure and content. An analysis of these materials and their corresponding questions and directives was undertaken to identify specific variables that relate to the constructs underlying subjects' performance on these tasks and to determine the extent to which those variables accounted for the variance associated with performance on the set of tasks. An important issue of validity was addressed by examining the extent to which the variables accounted for variance not only for the total population but for particular subgroups as well.

A first step in carrying out these analyses was to develop a grammar to describe the diversity of documents in the NAEP assessment. This grammar provided the basis for constructing Materials, Materials-by-Task, and Process variables. The grammar and the resulting variables are described in the following sections.

Procedures

A Grammar of Documents

Several grammars have been reported in the literature, e.g., Frederiksen's (1975), Kieras and Dechert's (1985), Kintsch's (1977), and Meyer's (1975), but each tends to apply to a particular type of discourse structure and genre (see de Beaugrande, 1981, and Mosenthal, 1985, for further discussion). The intent here was to devise a propositional grammar that

systematically describes the structure and content of diverse printed materials. The new grammar (Mosenthal & Kirsch, in preparation) was based on a significant revision of Mosenthal's (1985) taxonomic grammar of the expository continuum. In contrast with previous grammars, the grammar developed for this study was designed to be broad enough to account for documents as well as for other types of written discourse, such as technical writing, prose, and advertisements.

The grammar consists of 23 semantic-relation categories (see Appendix A for a listing). These categories are a synthesis from other propositional grammars. Information is hierarchically ordered into three basic levels: (1) the semantic feature, (2) the specific, and (3) the organizing category.

To illustrate these three levels, consider the Medicine Label document and its corresponding linguistic representation in Figure 1. Semantic features, the smallest unit of analysis, consist of arguments and relational terms. Arguments are typically nouns. For instance, in Figure 1 (following page), dosage in line 4, adults in Line 6, and teaspoons in Line 7 are arguments in the phrase "Dosage: Adults - 2 teaspoons every 4 hours." Relational terms qualify arguments and the relationships between arguments; such terms tend to be either verbs, adjectives, adverbs, and the like. For example, 2 in Line 8 is a relational term that stands in an ATTRIBUTE (ATT) relation with the noun teaspoons. The implied verb take (signalled by an *) in Line 5 relates adults as an AGENT (AG) in Line 6 to teaspoons as an OBJECT (OBJ) in Line 7.

The semantic features, including the verb, make up a unit of SPECIFIC (SPE) information, the next unit of analysis. In such instances, each SPECIFIC (SPE) makes up a micro proposition (or main feature) of document information. Note from Figure 1 (following page) that SPEs themselves can take on different semantic relation values, depending upon how they relate to preceding, contiguous information. For example, the SPE, directed by a physician (Lines 26 and 27), stands in a CONDITIONAL (COND) relation, signaled by unless, to the preceding SPE, do not exceed recommended dosage (Lines 20 to 24).

SPEs can serve to provide more specific information for a preceding SPE, as in the above example, or provide more specific information to a directly preceding ORGANIZING CATEGORY. An ORGANIZING CATEGORY (OC), the highest unit of analysis, consists of a generalized term or category that serves to summarize or synthesize more specific information (i.e., SPEs). The SPEs that make up an OC share a similar semantic feature or set of features represented by the OC.

For example, in the Medicine Label shown in Figure 1, there are two explicit OCs, dosage (Line 4) and caution (Line 20). The Medicine Label also contains a third, implicit OC, purpose of the medicine (Line 1). Implied categories consist of information that must be supplied by the reader and are denoted in the grammar by an (*). Note that this implied OC, although not directly stated in the document, is a common category shared by the SPEs, For Stuffed Noses (Line 2) and For Running Noses (Line 3). Also observe that there are two SPEs under the OC of dosage--namely, adults (Line 6) and

For Stuffed and Running Noses:

Dosage: Adults - 2 teaspoons every 4 hours; children over 6 years - 1 teaspoon every 4 hours.

Caution: Unless directed by physician, do not exceed recommended dosage. If drowsiness occurs, do not drive or operate dangerous machinery. Individuals with high blood pressure, heart disease, diabetes, or thyroid disease should use only as directed by a physician.

```
1      *|\OC purpose
2       |      |\SPE For Stuffed Noses
3       |  AND \SPE For Running Noses
4   *AND|\OC Dosage
5       |      |\SPE *take
6       |      |      |\AG  Adults
7       |      |      |\OBJ teaspoons
8       |      |      |       \ATT 2
9       |      |       \TEMP hours
10      |      |            |\ATT 4
11      |      |             \ATT every
12      |   AND \SPE *take
13      |              |\AG children
14      |              |       \ATT over six
15      |              |\ATT  teaspoon
16      |              |       \ATT 1
17      |               \TEMP hours
18      |                    |\ATT 4
19      |                     \ATT every
20  *AND \OC caution
21              |\SPE do exceed
22              |     *|\AG you
23              |      |\OBJ dosage
24              |      |       \ATT recommended
25              |      |\NEG not
26     unless   |  COND \SPE directed
27              |              |\AGT by physician
28              |             * \OBJ you
29       *AND (\SPE do drive
30         OR (\SPE do operate
31              |    *|\AG you
32              |     |\OBJ machinery
33              |     |       \ATT dangerous
34              |     |\NEG not
35           If|COND \SPE occurs
36              |              \AG drowsiness
37       *AND   \SPE should use
38                     |\AG individuals with
39                     |  *OR|\ATT blood pressure
40                     |     |       \ATT high
41                     |  *OR|\ATT heart disease
42                     |     |       *\ATT high
43                     |  *OR|\ATT diabetes
44                     |     |       *\ATT high
45                     |   OR \ATT thyroid disease
46                     |              *\ATT high
47             as COND \SPE directed
48                             |\MAN only
49                              \AG by physician
```

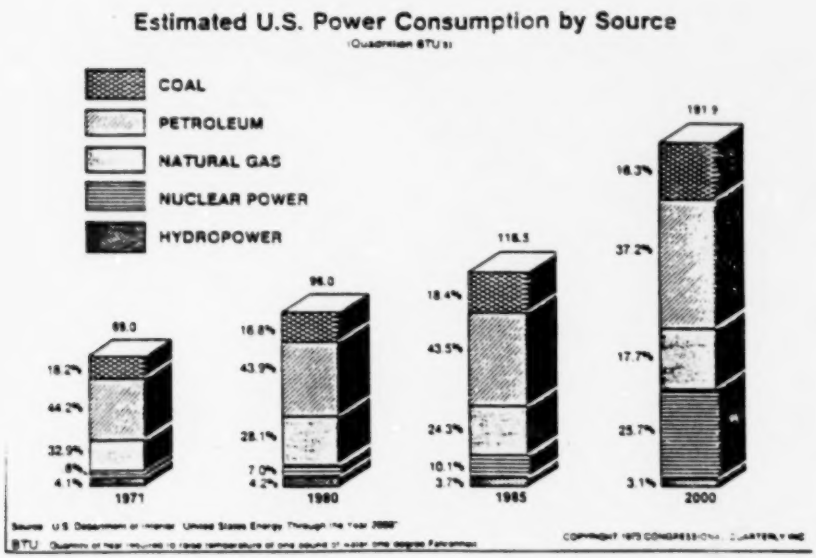
Fig. 1. A parsing of the Medicine Label document.

children (Line 13). Each of these SPEs, in turn, consists of a set of semantic features that stand in either an ATTRIBUTION (ATT) or TEMPORAL (TEMP) relation to the two SPE categories.

The OCs, SPEs, and semantic features taken together make up an information hierarchy. The macro structure of this hierarchy is defined by the relationship among the different OCs. Note that the OC structure for the Medicine Label consists of a single vertical structure (The OC, purpose, is related to the OC, caution, vertically by AND), as does the OC structure for the bar graph in Table 1 and Figure 2.

Table 1

The NAEP Power Consumption Bar Graph by Energy Source and Year



BEST COPY AVAILABLE

```
1  |\GC Estimated *(Amount of) US Power Consumption by Source
2  |     *(by Date)
3 *|\GC Comment
4  |       \ATT Quadrillion BTU's
5  |             \EQUI BTU
6  |                   \SPE quantity of
7  |                     |\ATT heat
8  |                     \SPE required
9  |                         *|\AG SPE 1
10 |                           \SPE to raise
11 |                               *|\AG quantity
12 |                                 \OBJ temperature of
13 |                                     |\ATT pound of
14 |                                     |     |\ATT one
15 |                                     |     \ATT water
16 |                                     \ATT degree
17 |                                         |\ATT one
18 |                                         \ATT Fahrenheit
19*|\GC Source
20 |   LOC |\SPE US Department of Interior
21 |       \SPE "United States Energy through the Year 2000"
22*|\GC Source of power
23 |     |\GC coal
24 |     |     \EQUI [     ]
25 |     |\GC petroleum
26 |     |     \EQUI [     ]
27 |     |\GC natural gas
28 |     |     \EQUI [     ]
29 |     |\GC nuclear power
30 |     |     \EQUI [     ]
31 |     |\GC hydropower
32 |     |     \EQUI [     ]
33 |\GC 1971
34*|\GC source of power
35 |   |-|\SPE [     ]
36 |   | |     *|\EQUI coal
37 |   | |      \ATT 18.1%
38 |   |-|\SPE [     ]
39 |   | |     *|\EQUI petroleum
40 |   | |      \ATT 44.3%
41 |   |-|\SPE [     ]
42 |   | |     *|\EQUI natural gas
43 |   | |      \ATT 32.9%
44 |   |-|\SPE [     ]
45 |   | |     *|\EQUI nuclear power
46 |   | |      \ATT  .6%
47 |   |- \SPE [     ]
48 |   |     *|\EQUI hydropower
49 |   |      \ATT  4.1%
50 |   \CI 69.0 *(quadrillion)
51 |             *\ATT BTU's
52 |\GC 1980
53*|\GC source of power
54 |   |-|\SPE [     ]
55 |   | |     *|\EQUI coal
56 |   | |      \ATT 16.8%
57 |   |-|\SPE [     ]
58 |   | |     *|\EQUI petroleum
59 |   | |      \ATT 43.5%
60 |   |-|\SPE [     ]
61 |   | |     *|\EQUI natural gas
62 |   |- \SPE [     ]
63 |   |     *|\EQUI hydropower
64 |   |      \ATT  4.2%
65 |   \CI 98.6 *(quadrillion)
66 |             *\ATT BTU's
67 |\GC 1985
68*|\GC source of power
69 |   |-|\SPE [     ]
70 |   | |     *|\EQUI coal
71 |   | |      \ATT 18.4%
72 |   |-|\SPE [     ]
73 |   | |     *|\EQUI petroleum
74 |   | |      \ATT 43.5%
75 |   |-|\SPE [     ]
76 |   | |     *|\EQUI natural gas
77 |   | |      \ATT 24.3%
78 |   |-|\SPE [     ]
79 |   | |     *|\EQUI nuclear power
80 |   | |      \ATT  10.1%
81 |   |- \SPE [     ]
82 |   |     *|\EQUI hydropower
83 |   |      \ATT  3.7%
84 |   \CI 116.5 *(quadrillion)
85 |             *\ATT BTU's
86 |\GC 2000
87*|\GC source of power
88 |   |-|\SPE [     ]
89 |   | |     *|\EQUI coal
90 |   | |      \ATT 16.3%
91 |   |-|\SPE [     ]
92 |   | |     *|\EQUI petroleum
93 |   | |      \ATT 37.3%
94 |   |-|\SPE [     ]
95 |   | |     *|\EQUI natural gas
96 |   | |      \ATT 17.7%
97 |   |-|\SPE [     ]
98 |   | |     *|\EQUI nuclear power
99 |   | |      \ATT  28.7%
100|   |- \SPE [     ]
101|   |     *|\EQUI hydropower
102|   |      \ATT  3.1%
103|   \CI 191.9 *(quadrillion)
104|             *\ATT BTU's
```

Fig. 2. Parsing of the NAEP Power Consumption Graph.

The macro structure is also defined by the relationship among the different SPE categories relative to OCs and to one another. Note in the Power Consumption graph, the major OCs consist of the years, 1971, 1980, 1985, and 2000 (Lines 33 to 52), in which different amounts of power are predicted to be consumed (Note that the copyright date of this graph was 1973). Another OC is the implicit category, source of power (see Lines 22, 34, 53, 68, and 87). The SPEs (Lines 23, 25, 27, 29, and 31) are the different bar legends that represent five different sources of power, i.e., Coal, Petroleum, Natural Gas, Nuclear Power, and Hydropower. The CONSTITUENCY IDENTIFICATIONs (or CIs) in Figure 2 (e.g., Lines 50 and 65) represent the total amount of power

produced in that year. Together, these OCs and CIs, with their SPEs and the ATTRIBUTES (ATTS) of these SPEs, make up the information hierarchy of the Power Consumption graph.

Defining the Materials Variables

Several studies (e.g., Atwood et al., 1984; Duffy, 1985) have shown that the complexity of materials influences the ability of users to process documents. In trying to account for complexity, this study identified six Materials variables that were derived from the grammar and that relate to length and organizational complexity. These variables included:

(1) the Total Number of OCs[1];

(2) the Number of Embedded OCs;

(3) the Deepest Level of Embedding for an OC;

(4) the Total Number of SPEs;

(5) the Number of Embedded SPEs; and

(6) the Deepest Level of Embedding for a SPE.

The Total Number of OCs and SPEs. The Total Number of OCs and SPEs was an arithmetic sum of the instances of each occurrence. For example, in the Medicine Label document (Figure 1), there are two explicit OCs compared to nine in the Power Consumption graph[2] (Figure 2). Similarly, there are ten SPEs in the Medicine Label document while the Power Consumption graph contains 28. Across the complete set of documents, the number of OCs ranged from 0 to 156. The number of SPEs ranged from six to 990.

The Total Number of Embedded OCs and SPEs. The Number of Embedded OCs consisted of all those explicit OCs that were embedded under an explicit OC.

Since there were no explicit OCs embedded off an explicit OC in either the Medicine Label or the Check Ledger, a score of zero was entered for this category for both documents.

Similarly, the Number of Embedded SPEs included all those SPEs that were embedded directly under a preceding, explicitly stated SPE. For example, in the Medicine Label document, we see three embedded SPEs, i.e., the three SPEs related conditionally (COND) to three preceding SPEs. In the Power Consumption graph, there are two embedded SPEs, namely *amount required (Lines 8 and 9) and *quantity to raise the temperature of one pound of water one degree Fahrenheit (Lines 10 to 18). These two SPEs are embedded under the SPE, quantity of heat (Lines 6 and 7).

The Deepest Level of OC and SPE Embedding. To determine the deepest level of OC embedding, we simply counted the number of levels that an explicit OC or CI, or series of explicit OCs and CIs, were embedded under a higher explicit OC or CI node. In the Medicine Label and the Power Consumption graph, we see that no OCs are embedded under a higher, explicit OC. Hence, the level of OC embedding for these two documents was scored a one, which represented the first level of information in the information hierarchy. Note particularly in the Power Consumption graph that the CIs are embedded under inferred OCs, namely, source of power; hence, these CIs did not count as being embedded.

Again, the deepest level of embeddings for SPEs consisted of the highest number of levels that an explicit SPE, or series of SPEs, were embedded. In the case of the Medicine Label document, we find three SPEs at the first level

of embedding. Consequently, this document received a score of 2 for the category of level of SPE embedding. This score represents the second level of information in the information hierarchy. In the Power Consumption graph, no SPEs are embedded under other SPEs, so it received a score of 0 for the category of level of SPE embedding.

Defining Materials-by-Task Variables

Another set of variables was identified focusing on the relationship between the document materials and their concomitant questions or directives. Research (e.g., Bransford, Franks, Morris, & Stein, 1979; Embretson, 1983; Kirsch & Guthrie, 1980; Morris, Bransford, & Franks, 1977) on the identification of cognitive components to account for task variance in reading comprehension has shown that models incorporating this interaction account for significantly more variance than those restricted to materials variables alone. Such component models acknowledge the fact that questions and directives set the goal, or purpose, for the reader or user. Moreover, these questions and directives help to determine what information should be processed and, thus, influence the strategies that are selected so that an information processing goal is met (Barnard et al., 1979; Kirsch & Guthrie, 1984b; Pearson & Johnson, 1978; Sticht, 1977).

The Number of OCs and SPEs necessary to complete a task. Again, based on the grammar used to parse the document tasks, the first two Materials-by-Task variables considered were the number of explicit OCs and SPEs, respectively, that a respondent had to process in order to complete a task correctly. For example, one question, based on the Power Consumption

graph in Table 1 and Figure 2, asked, "Which energy source is predicted to supply the MOST power in the years 1980, 1985, and 2000?" This question required readers to identify three OCs: 1980, 1985, and 2000. This task also required respondents to compare and contrast the percentage ATTRIBUTES of the five SPEs, or five Sources of Power, three times.

Another example taken from the Medicine Label document is shown in Figure 1. A question that asked, "What dosage should be given to adults?" requires the processing of one OC, dosage, and one SPE, *take \AG Adults \OBJ teaspoons \ATT 2, to arrive at the answer, 2 teaspoons.

The Deepest Level of OC or SPE Embedding Required by a Task. Two additional Materials-by-Task variables considered the deepest level of embedding at which a Materials-by-Task OC or SPE was located in the source material. It was reasoned that the more deeply an OC or SPE in a question or directive was embedded in the document macro structure, the more difficult it would be to locate and identify relevant information and, consequently, to complete a task (Meyer, 1985).

For example, referring back to the Power Consumption graph (Fig. 2), we might have asked, "What does BTU stand for in the graph's title caption, 'Quadrillion BTU's'?" To answer this, respondents would be required to process three SPEs, quantity of heat, *amount required, and *quantity to raise temperature of one pound of water one degree Fahrenheit. Because the third SPE appeared at the third level of embedding, this task would be judged to require processing at the third level of SPE embedding.

The levels of Materials-by-Task OC embeddings ranged from zero (where there were no explicit OCs in the materials) to three (which occurred in several tasks using a Bus Schedule, as well as in a task requesting particular information about a school's bus scheduling, lunch, and grading programs). Because the levels of SPE embeddings for the Materials-by-Task variable were almost entirely at the first level (and thus did not contribute to any score variance), this variable was eliminated from further analysis.

Defining Process Variables

In addition to recognizing the interaction that occurs between the document and the task, several studies (Fisher, 1981; Guthrie, 1988; Kirsch & Guthrie, 1984b; Kirsch & Jungeblut, 1986; Wright, 1980 a, b) have attempted to identify a set of procedures that describes the processes people engage in when performing tasks involving documents. These procedures can be summarized as follows: (1) Identify the given and requested information in a directive; (2) search the document until requested information is located; (3) make a match between the information identified in the document and the information requested in the document; and, (4) determine whether the match adequately meets the criterion of the task.

In light of these procedures, three variables were identified in this study as influencing the performance of users in completing document tasks. These included: (1) Degrees of Correspondence, (2) Type of Information, and (3) Degrees of Plausibility. These are considered in turn.

Degrees of Correspondence. This variable refers to the explicitness of the match between the information requested in the directive or question and

the corresponding information in the text. The reasoning here was that literal matches are more easily made than inferential text-based matches; and inferential text-based matches are more easily made than matches requiring specialized prior knowledge (Carpenter & Just, 1975; Hampton, 1987). The scoring scheme for Degrees of Correspondence is presented in Appendix B.

Type of Information. This variable refers both to the type and nature of information requested in the task. As previous research (e.g., Carpenter & Just, 1975) has shown, the fewer features needed to be identified and matched, the easier the task. Moreover, the fewer the restrictive conditions that must be held in mind in identifying and matching features, the easier the task. And finally, tasks that require literal matching of information contained in the document to the information requested in the question or directive are easier than those requiring some type of inference on elaboration process. The scoring scheme for Type of Information is presented in Appendix C.

Plausibility of Distractors. This variable refers to the extent to which information in the materials shares semantic information with the correct answer to a question or directive, but does not satisfy all conditions specified. In weighing alternative choices before completing a task, users must skim documents to identify which features, SPEs, or OCs best match the features of the information requested in a question or directive. As the number and relationships among plausible alternative choices increases, so does the difficulty of the document task (Brown, 1986; Drum, Calfee, & Cook, 1981). The scheme for scoring Plausibility of Distractors is presented in Appendix D.

Reliability

A final procedure involved obtaining reliability estimates for parsing the document tasks and scoring the process variables.

Scoring reliability for the Materials variables. To determine the reliability of parsing the document scale, 12 of the 37 documents were randomly selected. This was accomplished by listing each of the materials and selecting every third document. Next, a third person was trained to parse the tasks using the grammar developed. Reliability estimates were obtained by counting and comparing the total number of OCs and SPEs identified by the authors and the third person. The reason for estimating reliability based on these two levels and not specifically at the semantic feature level was because only the OC and SPE levels and general features within the SPE level were used to identify variables for this study; particular semantic feature relation categories per se did not define any variables identified in this study.

Among the Materials OC variables, there was 98 percent agreement between the authors and the third person in terms of identifying what constituted an explicit OC; 100 percent agreement for the number of OCs that were embedded; and, 98 percent agreement for the level of OC embedding. Similarly, among the Materials SPE variables, there was 96 percent agreement in terms of what information was identified as SPEs; 93 percent agreement for the number of SPEs that were embedded; and, 88 percent agreement for the level of SPE embedding.

Scoring reliability for the Materials-by-Task and Process variables. To determine the reliability for scoring the Process variables, the two authors first scored all the tasks in terms of the Degrees of Correspondence, Type of Information, and Plausibility of Distractors. Next, a list of tasks was devised by rank ordering the 61 tasks from easy to difficult based upon each task's P-value. Every third task was then selected from the list, so that a total of 21 tasks were identified representing the range of document task difficulty.

A third rater was then trained on the rules for each of the Materials-by-Tasks and for each of the Process variables. This training was carried out over tasks other than the 21 which also represented the range of document task difficulty.

Next, the third rater independently scored each of the 21 tasks on each of the Materials-by-Task and Process variables using the scoring schemes described in Appendices B, C, and D. When comparing this rater's scores to those of the authors', the following reliabilities based on percent agreement were achieved. On the Materials-by-Task variables, there was 96 percent agreement for the Number of OCs required by the question or directive; 99 percent agreement on the Level of OC Embedding required by the question or directive; and 90 percent agreement for the Number of SPEs requested in the question or directive. Among the Process variables, there was 95 percent agreement on Degrees of Correspondence; 86 percent agreement on Type of Information; and, 90 percent agreement on Plausibility of Distractors.

Results

In the previous section, three sets of variables were identified and described as they were hypothesized to contribute to our understanding of how these variables influence young adults' performance on document tasks representing different levels of difficulty. Here we describe the results of the analyses that were performed with these variables.

First, zero-order correlations were computed between scores on each of the Materials, Materials-by-Task, and Process variables and the percent correct or P-values for each of the document tasks. Percents correct represent the weighted percentages of 21- to 25-year olds who responded correctly to each of the 61 document tasks. The typically negative correlations (Table 2) between the materials and materials-by-task variables reflect the fact that higher P-values represent easier tasks. That is, more complex materials tend to be associated with more difficult tasks and, hence, tasks with lower P-values. It should be recalled that through the BIB spiraling design adopted for the NAEP assessment, each of the tasks was administered to a representative sample of some 1,500 young adults.

Zero Order Correlations

Table 2 shows the results of correlating scores on the three sets of variables with P-values. In addition to studying the pattern of relationships for the total population of young adults, we wanted to examine differences in the patterns of correlations across subgroups of interest.

Table 2

Zero Order Correlations among P-Values and Materials, Materials-by-Task, and Process Variables for Total Young Adult Population and Major Subgroups*

	Total P	White P	Black P	Hispanic P	0-8	9-12	High School Degree	Post High School Degree
Materials								
No. of OCs	-.38	-.37	-.38	-.43	-.36	-.40	-.39	-.29
No. of OCs Embedded	-.33	-.32	-.32	-.28	-.16	-.26	-.34	-.30
Levels of OC Embeddings	-.29	-.29	-.26	-.26	-.20	-.25	-.29	-.29
No. of SPEs	-.57	-.59	-.48	-.54	-.47	-.53	-.55	-.59
No. of SPE Embeddings	.01	.01	-.01	-.05	-.20	-.05	.03	.01
Levels of SPE Embeddings	.01	.02	-.05	-.01	-.12	-.11	.02	.06
Materials-by-Task								
No. of OCs	-.61	-.61	-.53	-.60	-.48	-.57	-.57	-.66
Levels of OC Embedding	-.25	-.26	-.18	-.22	-.19	-.23	-.23	-.52
No. of SPEs	-.52	-.53	-.50	-.53	-.45	-.51	-.50	-.52
Process								
Degrees of Correspondence	.49	.48	.49	.51	.36	.50	.51	.41
Type of Information	.85	.85	.83	83	.66	.80	.85	.80
Plausibility of Distractors	.59	.59	.57	.59	.43	.51	.60	.54

*Appendix E shows intercorrelations among these variables.

Therefore, correlations were also computed for White, Black, and Hispanic young adults, as well as for young adults having various levels of education: 0 to 8 years of schooling; 9 to 12, but no high school diploma; a high school diploma and/or some post-secondary experience; and at least a 2-year, post-secondary degree.

In general, the pattern of correlations found for the total population held for the subgroups as well. The highest correlation was found for Type of Information. This Process variable correlated at least .80 for each group,

with the exception of persons with 0 to 8 years of education (.66). Other variables having correlations above -.50 included one Materials variable (Number of SPEs), one Process variable (Degrees of Plausibility), and two Materials-by-Task variables (Number of Organizing Categories requested in the question or directive, and Number of SPEs asked for by the question or directive). Again, the correlations were lower for the group with 0 to 8 years of education. The lowest correlations were obtained for two of the Materials variables -- Number of SPEs Embedded and Level of SPEs Embeddings.

In addition to these general findings, there were several noteworthy patterns, particularly across levels of education. For example, with the Number of SPEs contained in the materials, the correlation rose from -.47 for those with 0 to 8 years of education to -.53 for those with 9 to 12 years, but no high school diploma; to -.55 for those having high school diplomas; to -.59 among those with at least a two-year degree. Conversely, the Number of SPE Embeddings in the materials correlated highest (-.20) for those with 0-8 years of education. It decreased to -.05 for those with 9-12 years; to .03 for those with high school diplomas; to .01 for those with at least two-year degrees.

Regression Analyses

Next, a series of regressions were run for the total population and for the major subgroups. The regressions were run using the multiple regression routine in the STATISTIX software package. The decision was made to enter into the regression equation those variables that had a zero-order correlation of at least .30. This level of correlation corresponds to roughly an alpha of

.01. Applying this rule resulted in the elimination of three Materials variables for each group: Level of OC Embedding, Number of SPE Embeddings, and Level of SPE Embeddings. Also, one Materials variable, number of OCs Embedded, was eliminated for Hispanics and for those with 0 to 8 and 9 to 12 years of education. In addition, one Materials variable (Number of OCs) was eliminated for those with the highest level of education, and one Materials-by-Task variable was eliminated for all groups (Level of OC Embedding).

Table 3 shows the results of the regression analyses. The numbers in the table represent the level of significance for each of the variables included in the model. Levels of significance at or below .05 represent variables in which the standardized regression weights are at least two times the standard error of measurement.

Table 3

The Significance Levels of the Predictive Variables in the Regressions for Total Population, Racial/Ethnic Groups, and Levels of Education with Overall R-Squared Values

	Total	Whites	Blacks	Hispanics	0-8	9-12	High School Degree	Post High School Degree
Materials								
No. of OCs	.66	.41	.70	.09	.33	.37	.70	--
No. of OCs Embedded	.13	.16	.23	--	--	--	.07	.40
No. of SPEs	.01*	.01*	.66	.29	.42	.22	.04*	.00*
Materials-by-Task								
No. of OCs	.01*	.01*	.05*	.00*	.08	.01*	.01*	.01*
No. of SPE	.01*	.01*	.03*	.01*	.09	.02*	.02*	.03*
Process								
Degrees of Correspondence	.01*	.01*	.01*	.01*	.21	.01*	.01*	.01*
Type of Information	.01*	.01*	.01*	.01*	.02*	.01*	.01*	.01*
Plausibility of Distractors	.27	.29	.38	.06	.94	.94	.12	.95
Multiple R	.94	.94	.90	.93	.75	.90	.94	.92
R-Squared	.89	.89	.81	.87	.56	.81	.88	.84
Adj. R-Sqd	.87	.88	.78	.86	.50	.78	.86	.82

* Significance levels at .05 and less.

Overall, there were five variables that contributed significantly to account for the variance in P-values across each of the groups of interest. Two of these were Process variables: Degrees of Correspondence and Type of Information. With one exception, these variables were significant at the .02 level or above for each group. The one exception was among young adults reporting 0 to 8 years of education. For this group, Degrees of Correspondence was not significant.

Two Materials-by-Task variables were also highly significant across each of the groups, with the exception again being among young adults reporting 0 to 8 years of education. These variables were Number of OCs and Number of SPEs stated in the question or directive that needed to be processed relative to the document in order to arrive at the correct answer.

The fifth significant variable, Number of SPEs, is a Materials variable that primarily reflects the amount of information contained in the document. Unlike the other four variables which were significant for each group (the exception, being among those with 0 to 8 years of education), this variable was significant for the total population; for White young adults; and for those reporting a high school diploma or a post-secondary degree. It did not reach significance, however, for Black or Hispanic young adults, or for those young adults who reported less educational experience than a high school diploma.

In addition to showing those variables that had regression weights that were at least two times the standard error, Table 3 also indicates the amount of variance that was accounted for by each of the regression analyses. For

the total population, the five significant variables accounted for 89 percent of the variance. Among racial/ethnic groups, the variance accounted for ranged from 89 percent for Whites, to 87 percent for Hispanics, to 81 percent for Blacks. Similarly, with one exception, these variables accounted for over 80 percent of the variance for young adults having different levels of education. The one exception was among those with 0-8 years of education. For this group, only 56 percent of the variance was accounted for.

Discussion

Summary of findings. The purpose of this paper was to identify the critical variables that underlie task difficulty for a nationally representative sample of young adults. To this end, three groups of variables were identified through the use of a propositional grammar and analyses conducted. These groups of variables related to: (1) materials, (2) the interaction between materials and task, and (3) process, or response, variables. Materials variables involved measures of document complexity and length. Materials-by-Task variables represented the type of information requested in a directive or question that needed to be processed in the materials in order to arrive at a correct answer. Process variables considered several factors that related to the processes of identifying, locating, matching, and responding to information in a document task.

The following variables were identified as significantly influencing document task difficulty for the total population of young adults, ages 21 to

25, as well as for select subgroups of this population. Among the five Materials variables investigated, the Number of SPECIFICs in the document was the only significant predictor of performance for the total population, for White young adults, as well as for respondents with a high school diploma and those with a post-secondary degree. The Number of SPECIFICs was a measure of length and amount of material in a document. As the Number of SPECIFICs increased, so did the difficulty of the document.

Among the four Materials-by-Task variables studied, two emerged as significant predictors for both the total population and the subgroups (with the exception of respondents with less-than-high-school-educational attainment level). These variables included the Number of ORGANIZING CATEGORIES and the Number of SPECIFICs needed to be processed in order to complete a task correctly. In general, as the Number of ORGANIZING CATEGORIES and SPECIFICs requested by a task increased, so did the difficulty level of a document task.

Finally, among the three Process variables identified in this study, two were significant predictors of both the total population's and the subgroups' performance. These included the Degrees of Correspondence and Type of Information. Degrees of Correspondence referred to how explicit the relationship was between the information requested in the task and the information in the materials; the less explicit this relationship was, the more difficult the document task was to complete.

Type of Information referred to the nature of processes that respondents engaged in to complete a task. Tasks that involved identifying, locating, and matching one set of features in the task to a corresponding set in the

materials were relatively easy compared to those that required identifying, locating, and contrasting several sets of features in the task to inferentially related sets of features in the materials. These latter tasks were substantially more difficult.

Significance of the findings. The significance of the above-mentioned variables is that, basically, they are excellent predictors of document difficulty for both the total population of young adults, as well as for subgroups of interest. For the total population, the sets of variables accounted for 89 percent of the variance in the distribution of percent-correct scores. Among racial/ethnic groups, the variance accounted for ranged from 89 percent for White, to 81 percent for Black, to 87 percent for Hispanic young adults. Among levels of education, the variance ranged from 56 percent for those with 0 to 8 years of schooling; to 81 percent for those with 9 to 12 years; to 88 percent for those reporting a high school diploma and/or some post-secondary experience; to 84 percent among those with at least a 2-year degree.

Given these results, this study makes two significant contributions. First, this study provides a theoretically based set of procedures for enhancing our understanding of document literacy. To date, testing a document's effectiveness has largely been based on an experimental design or qualitative analysis of users' performance on a single document task (Schumacher & Waller, 1985). In contrast, the current study illustrates an alternative approach, one based upon identifying the cognitive variables that profile task difficulty. For this set of procedures, task difficulty was

defined using the percentage of the population who were able to complete each task successfully within the domain of tasks. First, tasks were systematically represented using a descriptive grammar. Next, three sets of variables were generated that appeared to account for the rank ordering of tasks in terms of their difficulty. These sets of variables reflected the structure and complexity of the materials, the nature of the interaction between the materials and the corresponding question(s) and directive(s), and the processes of the reader.

Using this approach, similar task profiles could be established for any population of users and for any population of document tasks in any type of setting, e.g., military, government, or industry. This could be done by identifying the range of literacy tasks associated with any given occupation (Bond, Eastman, Gitomer, Glaser, & Lesgold, 1984). These tasks would then be administered to a sample that adequately reflected the population of users for that domain of tasks. The tasks would then be systematically parsed using a grammar, such as the one described in this paper. Next a set of critical variables, such as those generated for this study, would then be tested in terms of how well they described and predicted task difficulty.

A second contribution of this study is that by generating a set of variables that accounts for almost 90 percent of the variance over a broad set of document tasks, it provides a foundation for a theoretical model of document processing. Once available and validated, this model would provide a basis for: (1) identifying, developing, and refining instructional programs that systematically address the skills and strategies that readers use to

complete a wide variety of document tasks; and (2) a theory of document design to maximize the effectiveness of documents for all user groups in a document network. The need for such a model is particularly important given the functions documents serve in our lives and the increasing need for producing and processing documents more efficiently.

References

Athey, T. H., & Zmud, R. W. (1986). Introduction to computers and information systems. Glenview, IL: Scott, Foresman and Co.

Atwood, N., Baker, E., & Duffy, T. M. (1985). Cognitive approaches to the analysis of technical understanding. (CDC Technical Report No. 24). Pittsburgh, PA: Communications Design Center, Carnegie-Mellon University.

Barnard, P. J., Wright, P., & Wilcox, P. (1979). Effects of response instructions and question style on the ease of completing forms. Journal of Occupational Psychology, 52, 209-226.

Bassett, E. D., Goodman, D. G., Fosegan, J. S. (1981). Business records controls (5th ed.). Cincinnati, OH: South-Western Publishing.

Bond, L. Eastman, R., Gitomer, D., Glaser, R., & Lesgold, A. (1984). Cognitive task analysis of skills: Rationale and approach. Pittsburgh, PA: University of Pittsburgh, Learning and Development Center.

Bovair, S., & Kieras, D. E. (1981). A guide to propositional analysis for research on technical prose. (Technical Report No. 8). San Diego, CA; NPRDC.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 331-354). Hillsdale, NJ: Erlbaum.

Brown, J. (1986). Similarity relations in recognition: A model to explain Tulving's data. Journal of Memory and Language, 25, 197-201.

Burch, J. G., & Grudnitski, G. (1986). Information systems: Theory and practice (4th ed.). New York: Wiley.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. Psychological Review, 82, 45-73.

Chapanis, A. (1965). Words, words, words. Human Factors, 7, 1-17.

Charney, D. (1986). Redesigning and testing a work order form. (CDC Technical Report No. 28). Pittsburgh, PA: Communications Design Center, Carnegie-Mellon University.

Clark, H. H. (1972). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 12, 335-359.

Crandall, J. (1981, December). Functional literacy of clerical workers: Strategies for minimizing literacy demands and maximizing available information. Paper presented at the annual meeting of the American Association for Applied Linguistics, New York.

de Beaugrande (1980). Text, discourse, and process: Towards a multidisciplinary science of texts. Norwood, NJ: Ablex.

de Beaugrande, R. (1981). Design criteria for process models of reading. Reading Research Quarterly, 16, 261-309.

Department of Health and Social Security. (1983). The good forms guide. London: DHSS Forms Unit.

Drum, P. A., Calfee, R. C., & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. Reading Research Quarterly, 16, 486-514.

Duffy, T. M. (1981). Organizing and utilizing document design options. Information Design Journal, 2, 256-270.

Duffy, T. M. (1985). Readability formulas: What's the use? In T. M. Duffy & R. Waller (Eds.), Designing usable texts (pp. 113-143). Orlando, FL: Academic Press.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. Psychological Bulletin, 93, 179-197.

Felker, D. B., & Rose, A. M. (1981). The evaluation of a public document. (Technical Report No. 11). Washington, DC: Document Design Center, American Institutes for Research.

Firth, D. (1981). An investigation of the success of redesigned supplementary benefits documents. Information Design Journal, 2, 33-43.

Fisher, D. (1981). Functional literacy tests. A model of question-answering and an analysis of errors. Reading Research quarterly, 16, 418-448.

Frederiksen, C. (1975). Representing logical and semantic structure of knowledge acquired from discourse. Cognitive Psychology, 7, 371-458.

Guthrie, J. T., Seifert, M., & Kirsch, I. S. (1986). Effects of education, occupation, and setting on reading practices. American Educational Research Journal, 23, 151-160.

Guthrie, J. T. (1988). Locating information in documents: Examination of a cognitive model. Reading Research Quarterly, 18, 178-199.

BEST COPY AVAILABLE

Hampton, J. A. (1987). Inheritance of attributes in natural-concept conjunctions. Memory & Cognition, 15, 55-71.

Hartley, J. (1985). Designing instructional text. New York: Nichols.

Hartley, J., & Trueman, M. (1985). A research strategy for text designers: The role of headings. Instructional Science, 14, 99-155.

Janik, C. J., Hannah, J., Waney, S., Bond, S. J., & Hayes, J. R. (1981). Informed consent: Reality or illusion? Information Design Journal, 2, 197-207.

Jereb, B. (1986). Plain English on the plant floor. Visible Language, 20, 219-225.

Keller-Cohen, D. (1987). Organizational contexts and texts: The redesign of the midwest bell telephone bill. Discourse Processes, 10, 417-428.

Kieras, D. E., & Dechert, C. (1985). Rules for comprehensible technical prose: A survey of the psycholinguistic literature. (Technical Report No. 21, TR-85-ONR-21). Ann Arbor, MI: University of Michigan.

Kintsch, W. (1977). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

Kirsch, I. S., & Guthrie, J. T. (1980). The concept and measurement of functional literacy. Reading Research Quarterly, 19, 331-342.

Kirsch, I. S., & Guthrie, J. T. (1984a). Adult reading practices for work and leisure. Adult Education Quarterly, 34, 213-232.

Kirsch, I. S., & Guthrie, J. T. (1984b). Prose comprehension and text search as a function of reading volume. Reading Research Quarterly, 19, 331-342.

Kirsch, I. S., & Jungeblut, A. (1986). Literacy: Profiles of America's young adults. (Final Report No. 16-PL-02). Princeton, NJ: NAEP and Educational Testing Service.

Landesman, J. (1981). The "FISAP," before and after. The Journal of Business Communication, 18, 12-21.

Messick, S. (in press). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed.). New York: Macmillan.

Meyer, B. F. (1975). The organization of prose and its effects on memory. Amsterdam: North Holland.

Meyer, B. J. F. (1985). Signalling the structure of text. In D. H. Jonassen (Ed.), The technology of text (Vol.2, pp. 64-89). Englewood Cliffs, NJ: Educational Technology Publications.

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. Reading Research Quarterly, 17, 400-419.

Miller, R. B. (1984). Transaction and structures and format in form design. In R. Easterby & H. Zwaga (Eds.), Information design: The design and evaluation of signs and evaluation of printed materials (pp. 529-544). New York: Wiley.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519-535.

Mosenthal, P. B. (1985). Defining the expository discourse continuum: Towards a taxonomy of expository text types. Poetics, 14, 387-414.

Mosenthal, P. B., & Kirsch, I. S. (In preparation). Defining the structure of literacy materials: Towards a unified grammar of print. Princeton, NJ: Educational Testing Service.

Mosenthal, P. B., & Kirsch, I. S. (Submitted for publication). Designing effective adult literacy programs.

Odell, L., & Goswami, D. (1981). Writing in a non-academic setting. Research in the Teaching of English, 16, 201-224.

Pearson, P. D., & Johnson, D. D. (1978). Teaching reading comprehension. New York: Holt, Rinehart, & Winston.

Raines, G. (Ed.) (1980). Forms for the 80's: How to design and produce them. Philadelphia, PA: North American Publishing Co.

Rayner, D. (1982). Review of administrative forms report to the Prime Minister. London, England: Management & Personnel Office.

Schumacher, G. M., & Waller, R. (1985). Testing design alternatives: A comparison of procedures. In T. M. Duffy & R. Waller (Eds.), Designing usable texts (pp. 377-403). New York: Academic Press.

Scribner, S., & Cole, M. (1981). The psychology of literacy. Cambridge, MA: Harvard University Press.

Shilling, D. (1981). Plain English document design. Information Design Journal, 2, 244-250.

Sticht, T. (Ed.) (1975). Reading for working: A functional literacy anthology. Alexandria, VA: Human Resources Research Organization.

Sticht, T. (1977). Comprehending reading at work. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension (pp. 221-246). Hillsdale, NJ: Erlbaum.

Sticht, T. (1978). Literacy and vocational competency. (Occasional Paper No. 39, National Center for Research in Vocational Education). Columbus, OH: Ohio State University.

Sticht, T. (1982, January). Evaluation of the "reading potential" concept for marginally literate adults (Final Report FR-ET50-82-2). Alexandria, VA: Human Resources Research Organization.

United Kingdom. (1982). Administrative forms in government. London, England: CMND 8504, Her Majesty's Stationery Office.

Venezky, R. L. (1982). The origins of the present-day chasm between adult literacy needs and school literacy instruction. Visible Language, 16, 113-127.

Waller, R. (1984). Designing a government form: a case study. Information Design Journal, 4, 36-57.

Wright, P. (1978). Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. Instructional Science, 7, 249-312.

Wright, P. (1979). The quality control of document design. Information Design Journal, 1, 33-42.

Wright, P. (1980a). Strategy and tactics in the design of forms. Visible Language, 14, 151-193.

Wright, P. (1980b). Usability: The criterion for designing written information. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), Processing visible language (Vol. 2, pp. 183-205). New York: Plenum Press.

Wright, P. (1981). Informed design for forms. Information Design Journal, 2, 151-178.

Wright, P. (1984). Informed design for forms. In R. Easterby & H. Zwaga (Eds.), Information design and evaluation of signs and printed materials (pp. 545-577). New York: John Wiley & Sons.

Wright, P. (1988). Writing technical information. In E. Z. Rothkopf (Ed.), Review of research in education (Vol. 14, pp. 327-385). Washington, DC: American Educational Research Association.

Footnotes

1 Note that only explicitly occurring OCs were counted in (1), (2), and (3); inferred OCs were not counted because it was arbitrary to posit when an OC was or was not inferred.

2 CONSTITUENCY IDENTIFICATIONs (CIs) were also counted as OCs. This is because a CI acts much like an OC under a particular set of circumstances; in short, a CI relates two or more sets of quantitative categories which have been added, subtracted, multiplied, or divided to create a total.

Appendix A

Parsing Categories

ADVERSATIVE - Relates two assertions that make different claims about a cause-and-effect or conditional relationship.

ADV

E.g., **This study shows that the disease was caused by a virus; the other study suggests that the disease was caused by bacteria.**

AGENT - Relates an action with an immediate cause or outcome.

AG

E.g., **The man writes letters.**

AND - Relates two complementary actions or states.

AND

E.g., **The grass is green and the sky is blue.**

ATTRIBUTION - Relates an object or object set with an attribute or quality; applies directly only to SPECIFIC level information and not to RI or CI level information.

ATT

E.g., **The tall man walked quickly.**

CAUSE - Relates some cause, or influencing factor, with an effect or outcome.

CAUSE

E.g., **Martha quit the game because Matilda cheated.**

COMPARISON - Relates two statements that share the same feature or set of features.

COMP

E.g., **The virus in this jar is like the virus in the other jar; both cause influenza.**

BEST COPY AVAILABLE

CONCESSION - Relates two counterfactual or counterintuitive statements.

BUT E.g., Your aunt likes you but your uncle doesn't.

CONDITION - Relates a prerequisite existence of one thing to the

COND occurrence of something else.

E.g., If Martha goes, then Henry will go.

CONTRAST - Relates two topics, or statements, that vary in degree or kind

CONT in terms of one or more differentiating features.

E.g., John has more than Tim.

E.g., This solution is an acid; that solution is a base.

CONSTITUENCY IDENTIFICATION - Relates two or more sets of quantitative

CI categories which have been added, subtracted, multiplied, or divided.

E.g., Total Cost.

EQUIVALENT - Relates two identical objects or object sets.

EQUI E.g., John is the teacher.

Also relates a unique SPE to a given category.

E.g., Micky Bitsko's phone number is 335-2346.

EVIDENCE - General information is supported by subsequent facts or

EVI findings.

E.g., before a list of facts, one states:

Environment is a more important determinant of behavior than heredity. The following list of facts would constitute the evidence (EVI).

EXPLANATION - Previously presented facts or findings are stated in a more general manner, such as in the form of a law, principle, hypothesis, or prediction.
EXP
E.g., **after a list of facts, one concludes: These facts lead one to believe that environment is a more important determinant of behavior than heredity.**

GOAL - Relates a purposive state or event with a desired or intended outcome.
GOAL
E.g., **The man breaks the window to replace the cracked glass.**

LOCATIVE - Relates an object or action to some point in space.
LOC
E.g., **The dog is at home.**

MANNER - Relates a class of actions with an attribute that is a property of the action class. In short, describes the characteristics of actions.
MAN
E.g., **The dog ran quickly.**

NEGATIVE - Negates or nullified a state or action.
NEG
E.g., **John is not going, nor is his sister.**

OBJECT - Relates an action with an object that is affected by the action.
OBJ
E.g., **Bill hit John.**

OR - Relates two alternative actions or states.
OR
E.g., **Tasha is in New York or in London.**

ORGANIZING CATEGORY - A generalized term or category that serves to summarize or synthesize a specific or a series of specifics. Series of specifics are those that share a similar feature or set of features represented by the OC.

OC

E.g., **The category Balance in a check ledger.**

RECEIVER - A person, place, or thing that is a receiver of some action; a person, place, or thing that benefits from some action.

REC

E.g., **Mary gave John the book.**

SPECIFIC - More specific information about a topic, category or another specific is provided to conceptually qualify this topic, category, or specific. Often characterized by verbs or features refining an OC.

SPE

E.g., **After John went to the store, he went home.**

TEMPORAL - Relates a state or action to some point in time.

TEMP

E.g., **Nannette takes a long walk on Sundays.**

Appendix B

Scoring for Degrees of Correspondence

Score 5 if readers must complete the task by making a literal match between information (OC, SPE, or semantic features) in the directive or question and information in the text.

For example, consider the task to identify the date when a driver's license expires. This task involves making a literal match between date and expires in the directive with the OC, Expiration Date, on the driver's license. The SPE to this OC is the answer.

In a second example, a question on the document entitled, "Facts about Fire," might have asked, "What is the economic loss caused by industrial fires annually?" To answer this, readers might make the literal match between industrial fires in the question and industrial fires in the SPE, "Industrial fires in the U. S. each year cost some 46,000 firms about $236 million in property losses" to arrive at the correct answer, $236 million.

Score 4 if, to complete the task, readers must identify synonymous information in order to make a match between information (OC, SPE, or semantic features) in the directive or question and information in the materials.

For example, one task required readers to "Circle the sign that tells you that you are coming to a traffic light." To complete this, readers had to recognize that the information on the sign, Signal Ahead, in the materials was synonymous with coming to a traffic light in the directive.

Score 3 if, to complete the task, readers must make a low text-based inference in order to make a match between information requested in a directive or question and information in the materials.

For example, given an article entitled, "Facts about Fire," a question might have asked readers to identify, "For what sector of society do yearly fires have the severest economic consequences?" To answer this, readers would have to identify the cost of fires listed for (1) schools and colleges, (2) churches, and (3) industries; next, they would have had to compare the costs and identify $236 million as being the largest; finally, they would have had to recognize that this sum was associated with industries.

Score 2 if, to complete a task, readers have to make a high text-based inference in order to make a match between information requested in the directive or question and information in the materials.

For example, a question might ask, "On what page in the almanac can one convert Celsius to Fahrenheit?" In order to answer this question, readers might make the high text-based inference, temperature, as being the OC for Celsius and Fahrenheit. Using this OC, readers would then look to see if, in fact, that this was a category in the almanac under which the Celsius and Fahrenheit Conversion Table could be found as an SPE.

Score 1 if, to complete a task, readers must draw upon special prior knowledge in order to make a match between information in a directive or question and information in the materials.

For example, one task required readers to take information in an almanac table and re-enter it as the horizontal and vertical axes of a graph; in order to complete this task, readers were required to have unique knowledge of how to set up horizontal and vertical graph descriptors.

Appendix C

Scoring for Type of Information

Score 5 if, to complete a task, readers must:

a. Make a literal match between one SPE, or one or more features of an SPE (such as an AG, verb, or ATT), in the requested information in the directive or question and the same SPE or features in the materials; the answer identified as a consequence of this match is located in the same SPE or explicit OC node wherein the SPE or feature match was made.

For example, a task related to the Medicine Label document in Fig. 1 could have asked, "How many teaspoons of medicine should an adult take?" To answer this, readers must make a literal match between adult and teaspoons in the question to Adults (in Line 6) and teaspoons (in Line 7) in the materials. Note that the answer, 2 (in Line 8), appears under the same SPE node (i.e., *take in Line 5) as do Adults and teaspoons.

b. Make a literal match between one set of SPEs, which are supplied from a context other than from the directive or from the materials, to an OC in the materials that is stated or easily inferred.

For example, consider the task where one must fill in one's address on a job application. The OCs of the address include: Name, Apartment Number, Road, City, State, and Zip. These features could be considered separate SPEs, but because the same features are requested whenever one has to fill in one's address, they are instead grouped under one implied OC (i.e., *OC, Address). Thus, in this instance, the task is viewed as identifying one set of SPEs for the one *OC, Address.

Score 4 if, to complete the task, readers must:

a. Make a literal match between one SPE (or one or more features of an SPE) in the directive or question to a corresponding SPE (or feature in the materials) in the same SPE or OC in the materials; the answer identified as a consequence of this match is a low text-based inference that is located in the same or adjacent SPE or explicit OC node wherein the SPE or feature match was made.

For example, a task related to the Medicine Label document in Fig. 1 could have asked, "Before taking this medicine, what should you do if you have diabetes?" To answer this, the user must match diabetes in the question to diabetes (in Line 43). Based on this match, reader must then identify the SPE (in Line 37), should use, and the CONDITIONAL (COND) in Lines 47-49, as directed only by a physician. Based on this, readers must make the low text-based inference to answer the question, "I should first consult a physician." Again, note that the CONDITIONAL SPE was within the same SPE as the ATTRIBUTE, diabetes.

b. Identify in the materials an OC that represents the same kinds of SPEs (e.g., different sources of energy) but vary in degree (e.g., amount of yearly energy consumption by source) or salient characteristic; select the SPE that contains the key literal feature requested in the directive or question (e.g., year of production) by making a positive comparison.

For example, given a bar graph showing the amount of power (such as in Table 1) that different sources of energy produced in a four-year period, a question might ask, "Which energy source supplied the MOST power in the year

1980?" To answer this, respondents first have to identify the OC, 1980. Next, they have to select the energy source that has the highest percentage of BTU output in that year.

Score 3 if, to complete the task, readers must:

a. Infer an OC in the materials whose features are either the same as or synonymous with the requested information in the question or directive. Make a match between the inferred OC in the materials and the OC in the requested information. Once the match has been made, locate a SPE under the materials OC that completes the task directive or question correctly.

For example, a directive for a monthly telephone bill was to circle the month's charge for long distance calls. To complete this task, a respondent had to infer the OC, Calls Outside Local Area in the materials and identify this as being synonymous with the OC in the task, charge for long distance calls. After verifying this match, respondents then had to identify the SPE, $15.49, under the OC, Calls Outside Local Area.

b. First, identify an SPE in the question or directive within one OC in the materials. Second, identify a SPE in the question or directive within a second OC in the materials. Finally, conjoin the two SPEs in the materials in a way that satisfies the requested information. This applies primarily to information organized in a matrix fashion wherein the SPEs are organized under OCs that are arranged in a single horizontal column across the top of the matrix; the SPEs are contiguous on the same line.

For instance, a task had to do with the time that a person at a nursing home had returned. Information regarding each person was kept on a signout

sheet that had the following single horizontal column OCs: Date, Time Out, Resident, Returned, and Reason. To complete this task, a respondent had to first identify the OC, Resident, and then the SPE, Mrs. Farr. Next, the respondent had to identify the related OC, Returned, and its SPE. All these SPEs were on the same horizontal line in the document. Given that no time was specified in the SPE slot on the form, the answer was Mrs. Farr had not yet returned to her room.

c. **Identify an OC in the materials that represents the same kinds of SPEs (e.g., different sources of energy) but vary in degree (e.g., amount of yearly energy consumption by source) or salient characteristics. Select the SPE that contains the key selected feature requested in the directive or question (e.g., year of production) by making a negative contrast.**

For example, given a bar graph showing power consumption by source over a four year period, a question asked, "In the year 2000, which energy source is predicted to supply less power than coal?" To answer this, respondents first had to identify the SPE, coal, within the OC, 2000. Then they had to compare the amount of power consumption for each of the four energy sources relative to coal's power consumption amount.

Score 2 **if, to complete the task, readers must:**

a. **Identify and match an SPE or OC to a second OC several times such that each subsequent match depends upon the successful completion of the preceding match. This applies primarily to maps and to documents where the OCs are listed in both rows and columns and OCs are nested within OCs either in the horizontal or vertical column, or both.**

For example, given a street map of several city blocks, a question was as follows: "Begin at the Fourth Avenue side of City Hall; go west to Beech Street, north one block and then west two blocks. You are now on the corner of what two streets?" To complete this task, respondents had to locate a SPE on the map before identifying a next SPE; in the event that a preceding SPE was misidentified or mislocated, then the subsequent SPEs were incorrect.

Another example of this case involved the use of a bus schedule document. The question read, "When does the last bus from the Downtown Terminal leave Citadel?" To answer this, respondents first had to identify a series of interrelated OC's, representing different levels of embedding; these OCs were namely Outbound, Leave Downtown Terminal, Leave Citadel, and PM. The SPE within these nested OCs was the time, 6:45.

b. Identify in the materials the same SPEs in two different OCs of the same kind (e.g., two different dates) and compare the features of the SPEs to identify the appropriate SPE requested in the question or directive.

For example, given a bar graph of four different power sources over a four year period, a question asked, "In the year 2000, which energy source is predicted to supply a larger percentage of the total power than it did in 1971?" To answer this, respondents had to identify the two OCs, 2000, and 1971. Next, they had to compare the amounts of power produced by each of the five energy sources for these years. The SPE, Nuclear Power, had the ATT, .6%, in 1971 and 25.75% in 2000.

c. Create a SPE within an OC based upon observing the pattern of SPEs within a set of OCs in the materials.

For instance, a table showed the sales patterns by season for three years. A task was to predict the sales pattern for a season in an upcoming sales year. The set of OCs included the sales year and season within sales year. Another OC was amount of sales. Based upon the consistent pattern of SPEs of amount of sales over a three season period, respondents had to predict the SPE of amount of sales in a yet undetermined sales year.

d. Identify relevant features from a paragraph of specific information that precedes the question or directive and then enter these features under an OC with no SPEs specified.

For example, given a phone message form, a context was set where a caller, James Davidson, was inquiring whether a set of contracts were satisfactory; he requested an answer via return phone call by 2:00 pm. A task required respondents to complete the message form. To do this, respondents had to identify the OC, Message, and enter the relevant SPEs presented in the context.

Score 1 **if, to complete the task, readers must:**

a. Identify and match an SPE or OC to a second OC several times such that each subsequent match depends upon the successful completion of the preceding match; as a last step, select one SPE over another based upon SPE conditional information.

For example given a bus schedule, a question asked, "On Saturday morning, what time does the second bus arrive at the Downtown Terminal?" To answer

this, respondents had to identify the OCs, Arrive Downtown Terminal and AM. The second bus arrival time in the schedule was qualified with the SPE condition, Monday through Friday only. The third bus arrival time had no such condition; this time held for weekends as well as week days. Because it was Saturday, respondents had to keep in mind the conditional that special week day buses were not available on Saturday. Hence, the answer was the time specified for the bus listed third in the schedule.

b. First, identify and match an SPE or OC to a second OC several times on one information source such that each subsequent match depends upon the successful completion of the preceding match. Second, repeat this process for a second information source that has the same OCs but in a different format style than the first information source. Enter SPEs from the first information source under the corresponding OCs in the second information source.

For example, a task required respondents to identify the amount of oil exported by OPEC and NonOPEC nations to the United States over a six year period. This information was presented in a table in the World Almanac. Respondents then had to transfer this information from the almanac table to a graph, where they had to plot the amount of oil exported by OPEC and NonOPEC nations for four years.

Appendix D

Scoring for Degrees of Distractor Plausibility

Score 5 if there were no distractors.

This was characteristic of such things as the job application and social security form tasks where respondents had to enter SPEs under different OCs followed by blank lines.

For example, given the task to sign one's name to a Social Security card, there is only one OC, Signature, followed by a line for the person's signature; there are no other OCs followed by an empty blank in the Social Security card related to one's signature.

Score 4 if:

a. Distractors, like the answer, are exemplars of the same explicit or inferred OC and appear only within this OC. This was characteristic of such tasks that required respondents to identify information within a row or column of a table, graph, chart, or map.

For example, given a bar graph with five sources of power by year over a four year period, a task might require respondents to identify the energy source that supplied the MOST power in the year 1980. The answer was the SPE, petroleum. The other sources of power represented in the OC, 1980, would be distractors in the sense described in (a).

b. One or more distractors share a set of similar critical features as the answer but these distractors occur under a different OC node than the answer.

For example, a task asked respondents to identify the expiration date of a license. This date was specified under the OC, Expiration Date. In addition, another OC, Issue Date, specified a second date. The Issue Date specified was a distractor in the sense described in (b).

c. One or more OC distractors contain SPEs that share a set of critical features with the SPE under the OC answer. Applies primarily to documents arranged in a text format. Again, distractors are in a different OC node than the correct response.

For example, a task asked respondents to identify which of three buses (the bus numbers were the OCs) stopped at a particular intersection. Under the other bus numbers (i.e.,OCs), streets were also listed.

Score 3 if one or more distractors share a set of similar critical features as the answer with one distractor occurring within the same OC or node in which the answer appears. This applies to information organized primarily in text format.

An example of this was the Medicine Label task that asked, "How much medicine should be given to children over 7?" Note in Fig.1 under the OC, dosage (in Line 4), there are two amounts given (in Lines 8 and 10); both amounts are reported in the measurement of teaspoons and, hence, share this critical feature.

Score 2 if:

a. Under different OC categories the same exemplars are listed. Within an OC category, exemplars take on different quantitative attributes. Thus

distractors result not only from exemplars within an OC but also across OCs. Instances of this occur where comparisons are made between two or more OCs having the same list of exemplars.

For example, imagine a task that requires respondents to identify a source in a bar graph that shows one of five power sources increasing in output from one year to another. In both years, those power sources that decreased as well as the one that increased are identified; these decreasing sources serve as distractors as described in (a).

Score 1 if distractors are listed in a multiple choice format and the answer is a SPE within multiply embedded OCs. This information is generally presented in matrix form. Moreover, one of the distractors is an exemplar that occurs within the same level of OC embedding as the answer. Two other distractors occur within the same OC as the answer but at different levels of embedding.

An example of this was the multiple choice question, "When does the last bus from the Downtown Terminal leave Citadel?" The OC, Leave Citadel, appeared both under the OCs, Outbound from Terminal and Inbound toward Terminal. In addition, the OC, Leave Citadel was embedded under the OCs, AM and PM. The answer and a distractor appeared under the OC, Leave Citadel, that was embedded under the OCs, Outbound and PM. A second distractor appeared under the OC, Leave Citadel and the OCs, Inbound and PM. Finally, a third distractor appeared under the OC, Leave Citadel and AM.

Appendix E

Intercorrelations Among Materials, Materials-by-Task, and Process Variables

	Materials						M by T			Process		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Materials												
(1) No. of OCs	1.00	.25	.09	.52	.31	-.20	-.00	.13	-.10	-.45	-.40	-.19
(2) No. of OCs Embedded		1.00	.74	.18	-.18	-.05	.29	.41	-.04-	.02	-.29	-.34
(3) Levels of OC Embeddings			1.00	.12	-.12	.15	.41	.44	.03	.03	-.16	-.21
(4) No. of SPEs				1.00	.25	-.23	.31	.11	.17	-.15	-.53	-.39
(5) No. of SPE Embeddings					1.00	.26	-.15	-.13	-.02	.08	-.05	.00
(6) Levels of SPE Embeddings						1.00	-.13	-.17	.08	-.08	.09	.09
Materials by Task												
(7) No. of OCs							1.00	.50	.50	-.07	-.41	-.32
(8) Levels of OC Embeddings								1.00	-.06	-.03	-.22	-.21
(9) No. of SPEs									1.00	-.02	-.40	-.46
Process												
(10) Degrees of Correspondence										1.00	-.38	-.62
(11) Type of Information											1.00	-.03
(12) Plausibility of Distractor												1.00